J Geograph Syst (2004) 6:289–306
DOI: 10.1007/s10109-004-0138-z

Journal of Geographical Systems

Geospatial data resampling and resolution effects on watershed modeling: A case study using the agricultural non-point source pollution model

E. Lynn Usery, Michael P. Finn, Douglas J. Scheidt,[1] Sheila Ruhl, Thomas Beard,[2] and Morgan Bearden

U.S. Geological Survey, Mid-Continent Mapping Center, 1400 Independence Road, Rolla, Missouri 65401, USA (e-mail: usery@usgs.gov)

Received: 10 March 2003/Accepted: 21 May 2004

Abstract. Researchers have been coupling geographic information systems (GIS) data handling and processing capability to watershed and water-quality models for many years. This capability is suited for the development of databases appropriate for water modeling. However, it is rare for GIS to provide direct inputs to the models. To demonstrate the logical procedure of coupling GIS for model parameter extraction, we selected the Agricultural Non-Point Source (AGNPS) pollution model. Investigators can generate data layers at various resolutions and resample to pixel sizes to support models at particular scales. We developed databases of elevation, land cover, and soils at various resolutions in four watersheds. The ability to use multiresolution databases for the generation of model parameters is problematic for grid-based models. We used database development procedures and observed the effects of resolution and resampling on GIS input datasets and parameters generated from those inputs for AGNPS. Results indicate that elevation values at specific points compare favorably between 3- and 30-m raster datasets. Categorical data analysis indicates that land cover classes vary significantly. Derived parameters parallel the results of the base GIS datasets. Analysis of data resampled from 30-m to 60-, 120-, 210-, 240-, 480-, 960-, and 1920-m pixels indicates a general degradation of both elevation and land cover correlations as resolution decreases. Initial evaluation of model output values for soluble nitrogen and phosphorous indicates similar degradation with resolution.

---

[1]Now with SBC Communications, Inc., St. Louis, MO, USA
[2]Now with USGS, Mapping Partnership Office, Lafayette, LA, USA
Any use of trade, product, or firm names in this publication is for descriptive purposes only and does not imply endorsement by the U.S. Government.

We are grateful to Dr. David D. Bosch, ARS, Southeast Watershed Research Laboratory (SEWRL), Tifton, GA, for his assistance in installing and initializing the AGNPS model and, in particular, for supplying the specific land cover categorization developed by SEWRL for Little River and Piscola Creek.

Key words: GIS database development, watershed modeling, AGNPS, resolution effects, resampling effects

JEL Classification: C80, Q25, Q53

1 Introduction

Watershed models of volume, sediment load, quality, and peak flow depend heavily on geographic data sources, such as elevation, land cover, soils, and precipitation, which are commonly provided in geographic information systems (GIS). Researchers have been coupling GIS data handling and processing capability to water models for many years (Olivieri et al. 1991; Tim et al. 1992). However, it is still rare for GIS to provide direct inputs to the models. A few software developments - such as the GIS Weasel, a U.S. Geological Survey (USGS) computer program that interfaces GIS software with several water models in the Modular Modeling System (Leavesley et al. 2002; Viger et al. 2002) - are designed specifically to generate the parameters needed for operating specific models. The ability to use multiresolution databases for the generation of parameters is also problematic for distributed-parameter, grid based models, and the results of models operating under varying resolution conditions are still largely unexplored. Traditionally, these models are exhaustive, that is, they depend on data covering the entire region under study. This paper presents a discussion of database development procedures and the effects of resolution and resampling on GIS input datasets and parameters generated from those inputs for the Agricultural Non-Point Source (AGNPS) pollution model. By resampling, we mean that a data set at a resolution other than the one currently in use is to be analyzed.

The U.S. Department of Agriculture (USDA) developed AGNPS in response to the complex problem of managing non-point sources of pollution. AGNPS simulates the behavior of runoff, sediment, and nutrient transport from watersheds that have agriculture as their primary use. The model operates on a complete cell-by-cell basis and is a distributed parameter, event-based model. AGNPS requires 22 input parameters, including hydrologic data such as rainfall (amount and intensity), soils, drainage, agricultural management, and other information. The AGNPS model groups output parameters primarily by hydrology, sediment, and chemical content (Young et al. 1994, 1995; Witte et al. 1995).

Watershed models are commonly dependent on elevation for terrain morphology information and on land cover and soils for information on resistance to flow and chemical composition. These datasets are a part of the base data, which GIS are designed to process. Thus, the coupling of GIS software for parameter extraction for water models is a logical procedure. However, true integration of the modeling capabilities and GIS is rarely achieved because of the different data models used.

2 Watershed models and scale

Previous researchers have investigated various effects on a given region of spatial resampling and resolution on environmental model outputs. Vieux and Needham (1993) found that as grid-cell size increases model sediment yield increases by as much as 32% and that coarser resolution is the most important factor affecting these yields. Vieux (1993) assessed the effects of digital elevation model (DEM) aggregation by resampling a 30-m grid to 90, 150, and 210 m grids and found that as cell size increases flow-path length decreases (due to meander short-circuiting); area varies; and mean slope tends to become flatter (decreases).

For a given region, Garbrecht and Martz (1994) investigated the impact of DEM resolution on extracted drainage properties such as upstream drainage areas and channel lengths. They analyzed the effects of increasing DEM cell size ranging from 30 to 600 m (incremented in steps of 30 m). To compare hypothetical drainage network configuration with various drainage properties, they calculated a grid coefficient representing the ratio of a grid cell area to network reference area, which can be thought of as the ratio of the cell size to basin area. They found that for grid coefficients less than 0.01 (<1% of basin area), all extracted drainage properties are within 10% of the baseline reference values. For coefficients between 0.01 and 0.04, most of all drainage properties are within 10% of the reference values. For coefficients less than 0.08, all properties are within 20%, of the values. In addition, for coefficients greater than 0.08 properties increasingly diverge from the values. Finally, their analysis suggests that a DEM should have a grid cell area less than 5% of the basin area to reproduce drainage features with an approximate accuracy of 10%.

Wolock and Price (1994) used TOPMODEL to study the effects of topography on watershed hydrology. They showed that the map scale source of the DEM has an effect on model prediction of the depth to the water table, the ratio of overland flow to total flow, peak flow, and variance and skew of predicted stream flow. For example, the mean depth to the water table decreased with increasing coarseness of the data resolution and the maximum daily flow increased with increasing coarseness of the resolution. Hodgson (1995) demonstrated that the slope/aspect angle derived from the neighboring elevation points best depicts the surface orientation for a larger cell – either 1.6 times larger for the four nearest cells in a three by three window or 2.0 times larger for the eight nearest cells in a three by three window.

Because model prediction based on input datasets with low spatial resolution may not accurately reflect solute transport processes occurring *in situ*, Inskeep *et al.* (1996) compared observed and predicted data using two transport models with different levels of process description, and using model input parameters obtained from different resolutions. They showed that both models performed adequately with high-resolution model inputs (defined variably by data and “cases”) and that predictions using the Chemical Movement in Layered Soils (CMLS) model were less sensitive to data input resolution than the Leaching and Chemistry Estimation model, due in part to the fact that CMLS provides a simplified description of transport processes.

3 Objectives

Our development of AGNPS databases relies on a combination of procedures and macros to use commercial GIS software (specifically, ERDAS Imagine) to generate the parameters for AGNPS and a set of object-oriented code specifically designed to support this effort (Finn et al. 2002). This paper examines the problems of generating databases at multiple resolutions to support an analysis of the effects on the input data and the AGNPS parameters generated from those data. The paper thus provides an assessment of the effects of resolution and resampling. The next section of the paper discusses our approach to the creation of the databases. Section 5 presents the study areas and data sources, and Section 6 documents the methods used for database development. Section 7 provides a tabulation and comparison of statistical results, including differences in area tabulations and values at specific locations resulting from resolution differences. A final section draws conclusions from this work.

4 Approach

We used USGS 30-m DEMs from the National Elevation Dataset (NED) (Gesch et al. 2002; USGS 2002a) and land cover from the National Land Cover Data (NLCD) (Vogelmann et al. 2001; USGS 2002b) as the base for the extraction processes. The NLCD information was augmented with land cover extracted from Landsat Thematic Mapper (TM) data acquired in 1997 and 2001. In addition, we acquired high-resolution (3-m) elevation and land cover data to help determine resolution effects. We generated the soil databases from USDA soil surveys by scanning mylar separates of soil polygons, then rectifying, vectorizing, and tagging the resulting digital data. The soil data from vector format were resampled to the 30-m and 3-m base resolution grids to match land cover and elevation datasets. To assess the effects of resolution and resampling on model inputs, we resampled the 30-m raster data to 60-, 120-, 210- (approximately 10 acres, commonly used by the USDA), 240-, 480-, 960-, and 1,920-m cells.

5 Study areas and data sources

The four watersheds examined include Little River and Piscola Creek, Georgia; Sugar Creek, Indiana; and EL68D Wasteway, Washington (Table 1). Little

Table 1. Study areas

Watershed	Counties	Drainage	Hectares	Approximate center
Little River, Georgia	Turner, Worth, Tifton	Into the Withlacoochee River, then the Suwannee	44,414	30° 49′ 00″ N 83° 39′ 00″ W
Piscola Creek, Georgia	Brooks, Thomas	Into the Withlacoochee River, then the Suwannee	33,242	30° 46′ 00″ N 83° 38′ 00″ W
Sugar Creek, Indiana	Henry, Hancock, Madison	Into the Driftwood River, then the East Fork of the White	23,976	39° 55′ 00″ N 85° 43′ 00″ W
EL68D Wasteway, Washington	Adams, Franklin	Into the Potholes Canal, then the Scooteney Reservoir	37,719	46° 49′ 11″ N 119° 02′ 13″ W

River, Sugar Creek, and EL68D Wasteway were selected because they are National Water-Quality Assessment (NAWQA) Program sites where USGS personnel do periodic sampling. We added Piscola Creek because of previous work in that watershed and data availability, including 5 years of continuous water-quality samples at nine stations. In this paper we focus on results from the Little River and Piscola Creek, Georgia watersheds (Fig. 1).

5.1 Watershed boundaries

We used two sets of watershed boundaries for this study. The USGS collects and assesses information on water chemistry, hydrology, land use, and stream habitat in more than 50 major rivers across the nation as a part of the NAWQA Program and has established boundaries for these watersheds (Hamilton 2002). Part of the program is concerned with water quality and non-point sources in agricultural watersheds (Berndt et al. 1998; Gilliom et al. 2002); thus, one of the boundary sets used was the NAWQA watershed boundaries. Because the NAWQA boundary does not usually match the flow according to the DEMs, because of resolution and accuracy issues, we used a second set of watershed boundaries. This second set was extracted directly from the DEMs using the GIS Weasel (described earlier). Because the DEM determines the resulting GIS Weasel boundary, this boundary is consistent with slope and other data derived from the DEM. Table 2 contains a comparison of watershed areas resulting from the NAWQA and DEM-extracted boundaries at various resolutions for Little River, Georgia and a calculation of the grid coefficients based on Garbrecht and Martz (1994) for the Little River basin area. Figure 2 provides a comparison of the two boundaries at 30-m resolution.

Fig. 1. Georgia study areas are the Little River and Piscola Creek watersheds

Table 2. Comparison of watershed areas and grid coefficients[a]

Resolution (m)	NAWQA (ha)	GIS Weasel (ha)	Difference	Grid coefficient NAWQA	Grid coefficient Weasel
30	33423.8	34885.8	1462.0	0.0009	0.0009
60	33702.5	35089.2	1386.7	0.0018	0.0017
120	34076.2	35493.1	1416.9	0.0035	0.0034
210	34631.7	35986.1	1354.4	0.0061	0.0058
240	34859.5	36241.9	1382.4	0.0069	0.0066
480	36426.2	37739.5	1313.3	0.0132	0.0127
960	39444.5	40458.2	1013.7	0.0243	0.0237
1920	45711.4	46418.9	707.5	0.0420	0.0414

[a]As defined by Garbrecht and Martz (1994)

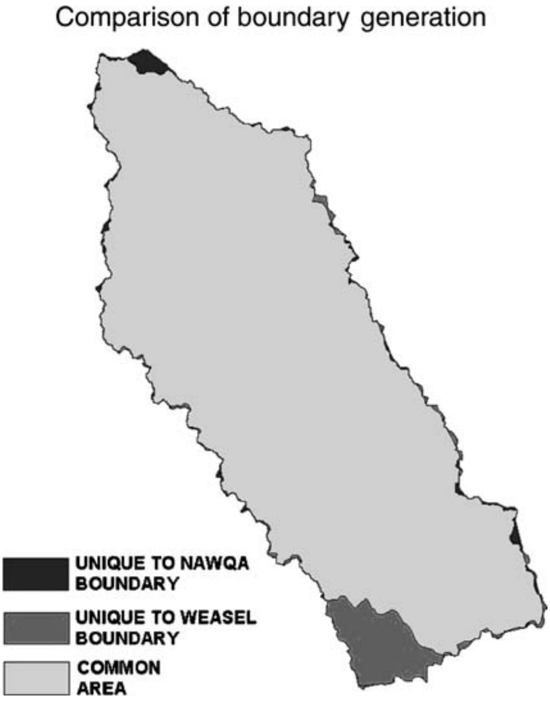


Fig. 2. Comparison of NAWQA and GIS Weasel watershed boundaries for Little River, Georgia at 30-m resolution

6 Methods of database development

6.1 Land cover

Two sources of land cover were developed for each watershed: NLCD with 30-m pixels and classification at 3-m pixels from high-resolution panchromatic images utilizing the Anderson classification scheme (Anderson *et al.* 1976). In addition, a classification at 30-m pixels from recent TM images supported by field reconnaissance was employed utilizing a specific scheme developed for Piscola Creek and Little River by the USDA Agricultural Research Service (ARS) (Table 3). Accuracies of classification for the 30-m and 3-m data match the NLCD at about 90 %, and absolute classification accuracy in the Little River watershed was computed to be 90 % by comparison with field-sampled sub watershed areas for which wall-to-wall categories are known.

6.2 Elevation

Elevations for the watersheds were acquired from the USGS NED. These data are resampled to a 30-m post spacing (pixel size) and achieve accuracies of ± 3-m vertical root-mean-square error (RMSE) in all watersheds. Higher resolution elevation data with 3-m post spacing were generated by a combination of automated correlation of high-resolution panchromatic stereo images and interactive editing and correction. Accuracies of 1-m vertical RMSE were attained in all watersheds for the high-resolution data.

6.3 Soils

Soil data were acquired directly from the USDA for Little River and Piscola Creek as vector polygons. For Sugar Creek and EL68D Wasteway, soil data were generated by raster scanning the black-line Mylar separates of the USDA soil surveys and converting these to vector polygons. The polygons were then tagged with attributes for soil types and other information from

Table 3. Land cover categories

NLCD classification categories	ARS specific land cover categories	Anderson land cover categories
Low-intensity residential	Urban	Urban (unclassified)
High-intensity residential	Pecan	Transportation/ Utilities
Commercial/ Industrial/ Transportation	Mature deciduous	Unknown orchards
Urban/ Recreational grasses	Mature planted pine	Mature deciduous
Orchards	Mixed deciduous / Pine	Mature pine
Deciduous forest	Young planted pine	Young pine
Evergreen forest	Water	Forest (unclassified)
Mixed forest	Wetland	Mixed forest
Shrub land	Forested wetland	Water (unclassified)
Open water	Crop	Reservoirs
Woody wetlands	Disturbed or harvested land	Wetlands
Herbaceous wetlands	Urban	Forested wetlands
Grasslands herbaceous	Pecan	Crops (unclassified)
Pasture/ Hay	Mature deciduous	Mixed grasses
Row crops		Corn
Small grain		Cotton
Transitional		Peanuts
Fallow		Unclassified
		Barren land
		Disturbed or harvested land

the soil surveys. The polygon data were resampled to 3- and 30-m grids to match the elevation and land cover datasets.

6.4 Precipitation

Precipitation data to support the AGNPS model were obtained from the National Weather Service in each of the watershed areas. This dataset is not a spatially distributed parameter for AGNPS. Since AGNPS is event based, a single value is applied to the entire watershed.

7 Resampling to coarser resolutions

The 3-m and 30-m datasets provide independent collections of input data for the parameter-generation process of AGNPS and allow examination of the effects of resolution on both input and output parameters. We also resampled the GIS datasets to determine the effects of resampling. In the case of land cover and soils, the resampling was performed with a nearest neighbor method consistent with the categorical scaling of the data. For elevation, we resampled using the bilinear interpolation method (ERDAS 1999, p 367) to achieve the lower data resolutions. From the resamplings of the 30-m data to 60-, 120-, 210-, 240-, 480-, 960-, and 1920-m cells, parameters were generated.

8 Results

Using the various watersheds, input data layers, and resolutions, our analysis can be divided into two areas: 1) effects of resolution, and 2) effects of resampling. We used independent collections of elevation and land cover to test the effects of resolution, and we resampled 30-m data for elevation, land cover, and soils to test the effects of resampling.

8.1 Effects of resolution

The independent data collections at 3 m and 30 m for elevation and land cover allow an assessment of the effects of resolution on elevation values at specific locations, on land cover types in specific locations, and on the resulting values of the input parameters for AGNPS. To test these effects, we selected 500 points randomly over each watershed and extracted the values at each point in both the 3- and 30-m data. Table 4 provides results for 20 representative points of the 500 points with corresponding 3- and 30-m elevation and land cover values for the Little River watershed. The 3-m land cover was recoded (generalized) from the Anderson classification to the ARS-specific classification to allow comparison.

Using the values over the 500 points for Little River, we determined the correlation between elevation values. Elevations resulted in a linear (Pearson product moment) correlation coefficient (r) of 0.90 between the 3- and 30-m resolution data. This value of r is remarkable considering that these were not resampled versions of the same dataset but were independent collections. We

Table 4. Sampling of points for land cover and elevation comparisons for Little River, GA, (land cover classification for the 3-m data were recoded from the Anderson classification to the ARS-specific classification to allow for comparison)

Easting	northing	3-m LC	30-m LC	3-m Elev	30-m Elev
239589	3504260	Crop	Mature planted pine	119	122
241209	3503180	Crop	Crop	125	124
256449	3486470	Urban	Crop	102	103
252039	3491360	Mature deciduous	Wetland	84	85
240369	3516350	Mixed deciduous/ Pine	Mature planted pine	132	132
253959	3486830	Urban	Crop	90	85
253539	3496400	Urban	Crop	111	111
246369	3497360	Mixed deciduous/ Pine	Wetland	95	94
247779	3512330	Urban	Urban	130	130
256179	3491270	Crop	Crop	97	97
244239	3498170	Mixed deciduous/ Pine	Mature planted pine	106	106
238449	3515090	Young planted pine	Mature planted pine	132	130
254589	3486920	Mature planted pine	Crop	84	85
244749	3504560	Crop	Crop	121	119
250929	3495140	Crop	Crop	107	100
247719	3498890	Crop	Crop	115	112
244359	3507260	Crop	Disturbed or harvested land	116	115
255579	3491240	Mixed deciduous/ Pine	Wetland	95	94
252339	3500660	Crop	Crop	113	115
247719	3508160	Crop	Crop	117	116

used multinomial regression to compare the land cover classification results and obtained no significant correlation. McFadden's coefficient of determination (pseudo R^2) value (Long 1997) for this comparison yielded a value of 0.139, which implies essentially no correlation. The percentage of the 500 random points that show the same land cover category at the 3 and 30 m resolutions is only 21.4, again indicating little correlation across resolutions. These results indicate that from two independent collections of elevation and land cover with an order of magnitude difference in linear resolution, common elevation values result at randomly sampled locations with no significant loss of information resulting from resolution difference, but land cover classes do not correspond across resolutions. This result may be a function of the resampling methods used: bilinear interpolation for elevation and nearest neighbor for land cover.

8.2 Effects of resampling

We evaluated resampling effects on the databases in a variety of ways, including comparing the change in area of specific land cover categories across resolution and comparison of the elevation values at specific locations in datasets of different resolution. Comparing the coarser resolution datasets shows the effects of resampling. Figure 3 shows the effects of resampling on raster resolution for a generalized land cover classification in the Little River watershed. Table 5 shows a comparison of the total areas of land cover in Little River when 30-m data are resampled to lower resolutions using the same generalized land cover classification as shown in Fig. 3. The values in the table are percentages of the 30-m land cover areas. This demonstrates that as the data are resampled to consecutively lower resolutions, the land cover categories decrease in accuracy (increase in variability) when measured as a function of the percentage of the 30-m categories. We obtained similar results in all watersheds.

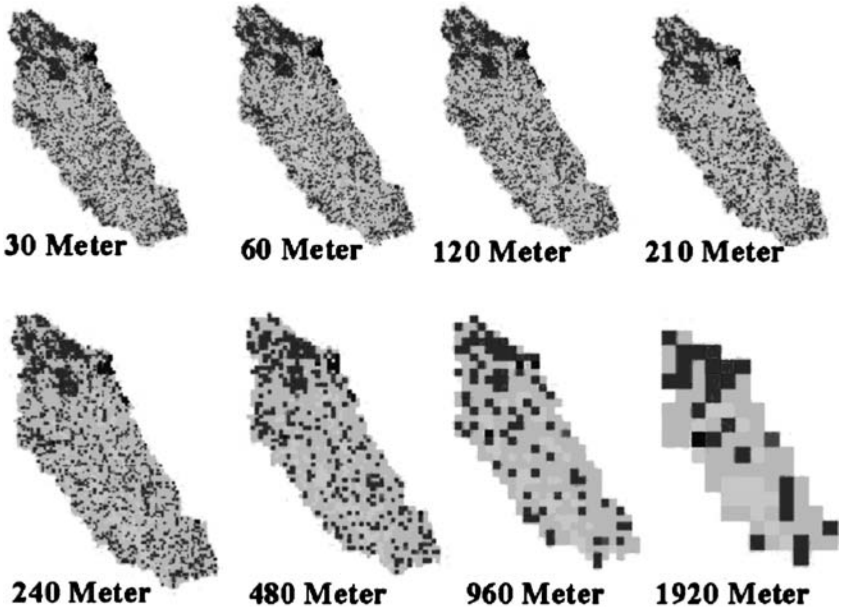


Fig. 3. Effects of resampling on raster resolution on a generalized land cover classification in the Little River, Georgia watershed

Table 5. Comparison of land cover values across resamplings for Little River. Values are percentages of 30-m land cover category areas. (Based on a generalized land cover classification)

	60 m	120 m	210 m	240 m	480 m	960 m	1,920 m
Water	100.12	94.50	121.23	97.04	98.56	65.71	0.00
Urban	104.04	94.34	100.28	76.73	68.51	35.97	89.98
Transitional	100.54	96.96	92.18	90.15	81.95	69.34	90.20
Deciduous	103.35	101.65	102.12	94.36	120.21	156.28	97.74
Pine	100.11	98.48	97.90	96.80	86.37	69.99	48.35
Mixed	98.18	101.29	94.74	98.41	86.33	65.95	148.47
Crop	98.99	97.72	95.42	95.80	91.40	88.33	74.46
Wetlands	99.50	100.27	99.98	102.78	101.91	105.47	82.45

Another method to compare the effects of resampling on data values is to compare a single profile of values across the watershed at various resampling sizes. Elevation values along the line across the Little River watershed depicted in Fig. 4 are shown in Fig. 5. As with the resolution analysis above, the profiles indicate good correspondence of elevation values with resampling; however, large down-sampling from 30 to 960 and 1,920 m degrades the data beyond practical use. This degradation is obvious from Fig. 5 and is consistent with a visual interpretation of resampling effects from Fig. 3.

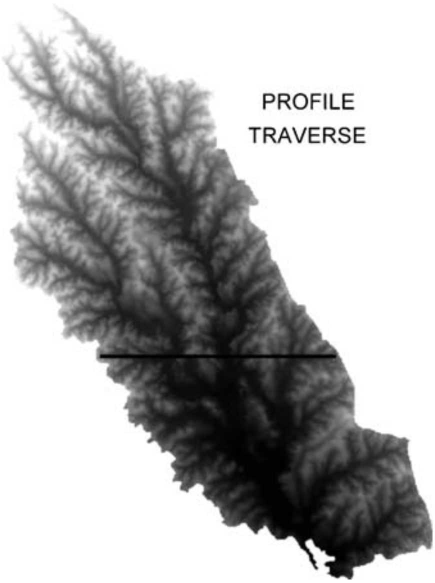


Fig. 4. Profile transverse across Little River watershed (elevation profiles are shown in Fig. 5)

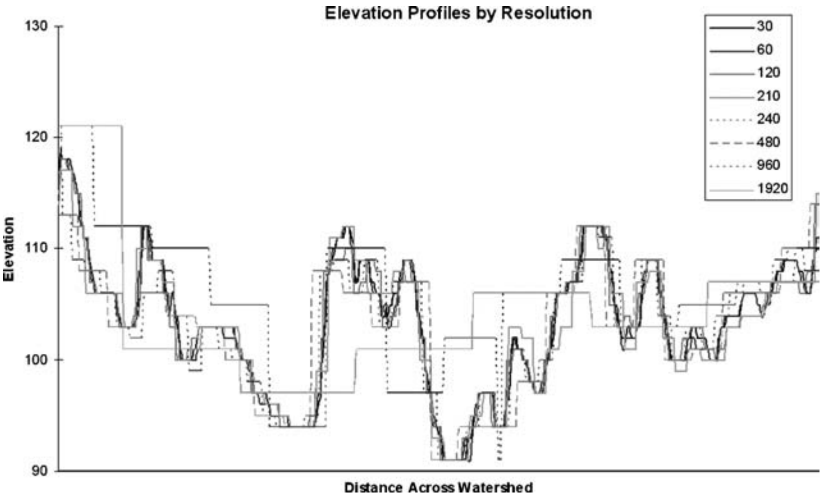


Fig. 5. Spatial resampling effects shown by an elevation profile

As in the resolution analysis, elevation values for each resampled cell size (60, 120, 210, etc.) over a 500-point random sample were regressed against the elevation value at the corresponding point in the original 30-m grid. Figure 6 shows the effects of resampling elevation data for Little River as measured by linear regression; specifically the r-value decreases with increasing pixel size. Figure 6 shows similar effects for Piscola Creek. Note in Fig. 6 (b), the lower accuracy for the 210 m compared with the 240 m. An understanding of the resampling methodology may explain this apparent anomaly from the trend since bilinear resampling uses four surrounding pixels; even numbers of four pixels (for example, 30 m to 60 m to 120 m to 240 m) should show consistent results. The 210 m is not an even multiple of four from the 30-m original data and thus yields lower accuracy than the 240 m.

In addition to analyzing the resampling of the GIS input datasets (elevation and land cover), we also examined the effects of resampling on derived input parameters for AGNPS. We limit our discussion to two example derived-parameters: land slope, a continuous variable, and flow direction, a discrete variable. The parameters were derived from the 30-m

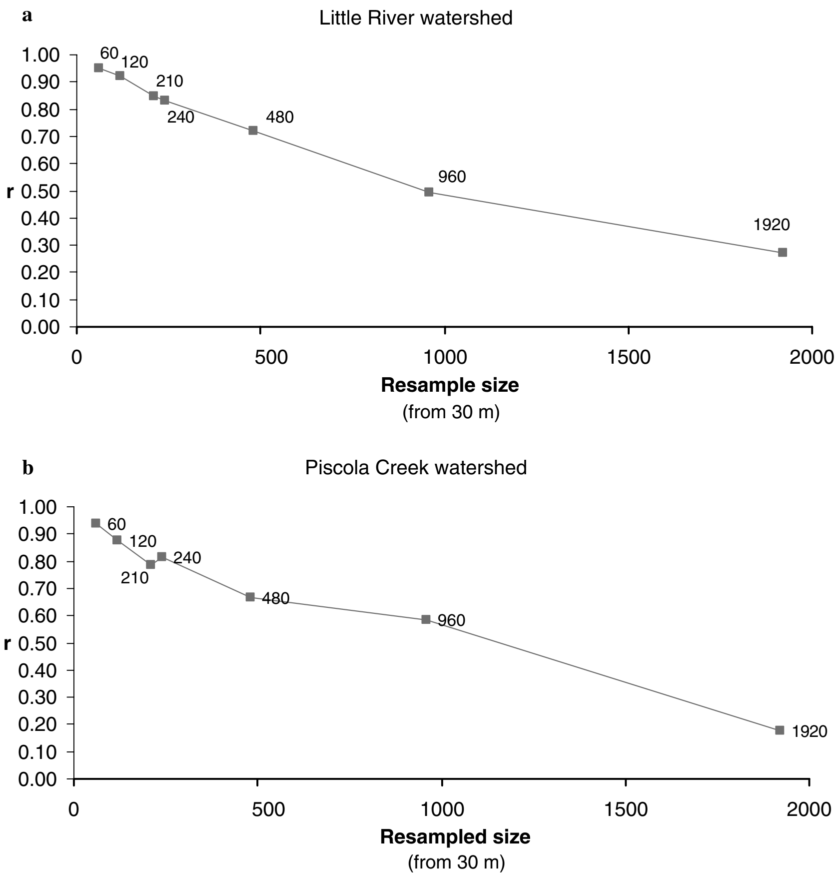


Fig. 6. Resample correlation coefficient (r) for elevation. **a** Little River, **b** Piscola Creek

elevation data and then resampled to the various resolutions. Figures 7 shows the effects of resampling the 30-m land slope data for various resolutions from Little River and Piscola Creek as measured by linear regression. For the random set of 500 test points, land slope derived from the 30-m data was compared with the corresponding value in the resampled data. The results show a continual degradation of r as cell sizes become larger. Figure 8 shows the effects of resampling on the flow direction data for Little River as measured by multinominal regression between 30-m and resampled datasets (McFadden's pseudo R^2 value). Figure 8 shows the same for Piscola Creek. Note that, in Fig. 8, the 210-m value reflects the effect of the bilinear resampling discussed in the previous paragraph. Figure 8 demonstrates that there is very little correlation (30- to 60-m case) essentially to zero correlation (30- to 1,920-m case) between categories after resampling of derived input parameters.

Preliminary investigation of resampling phenomena verses AGNPS model output parameters for two of 51 output parameters shows that there is very little to zero correlation between model output at various cell sizes compared

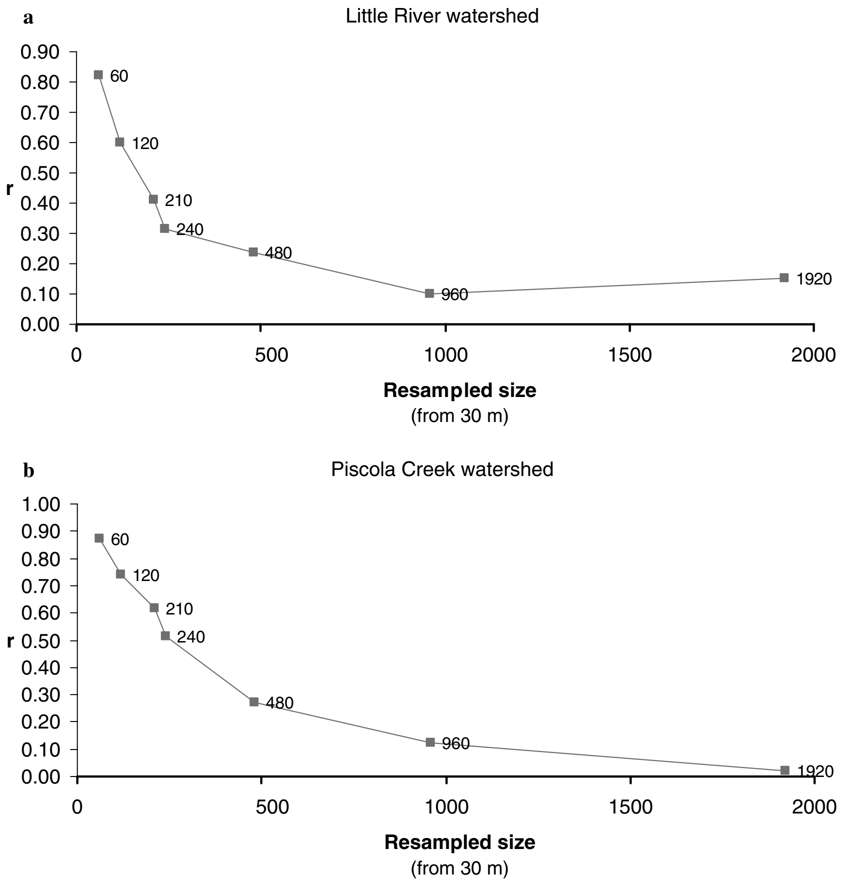
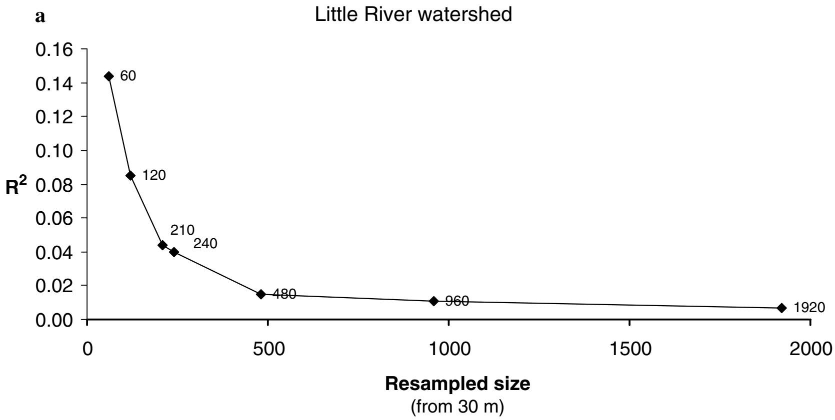


Fig. 7. Resample correlation coefficient (*r*) for elevation: **a** Little River, **b** Piscola Creek



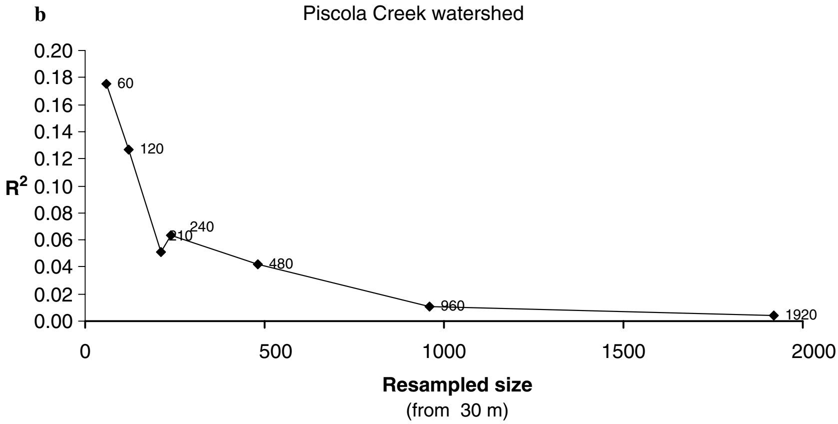


Fig. 8. Resample McFadden's coefficient of determination (pseudo R^2) for flow direction: **a** Little River, **b** Piscola Creek

to the corresponding 30-m output. For this preliminary investigation, we used soluble nitrogen and soluble phosphorous concentration (in parts per million) output parameters and calculated their correlation coefficient (r) over a 500-point random sample for Little River watershed (Fig. 9). This seems to attribute all the change in output parameters directly to the effects of resampling on the input parameters because no other parameters were modified. Further analysis of the output parameters as a function of the spatial input is the subject of future research.

This works corroborates Inskeep et al. (1996) who showed that model predictions based on input datasets with low spatial resolution may not accurately reflect physical processes occurring in the environment. Our findings generally corroborate Worlock's and Price's (1994) findings that data resolution has an effect on model prediction for a variety of watershed related parameters. The results of our investigation generally correlate to Garbrecht's and Martz' (1994) findings when calculating their dimensionless grid coefficient, representing the ratio of grid area to basin area, and that a grid area should be less than 5% of the basin area to reproduce accurate features.

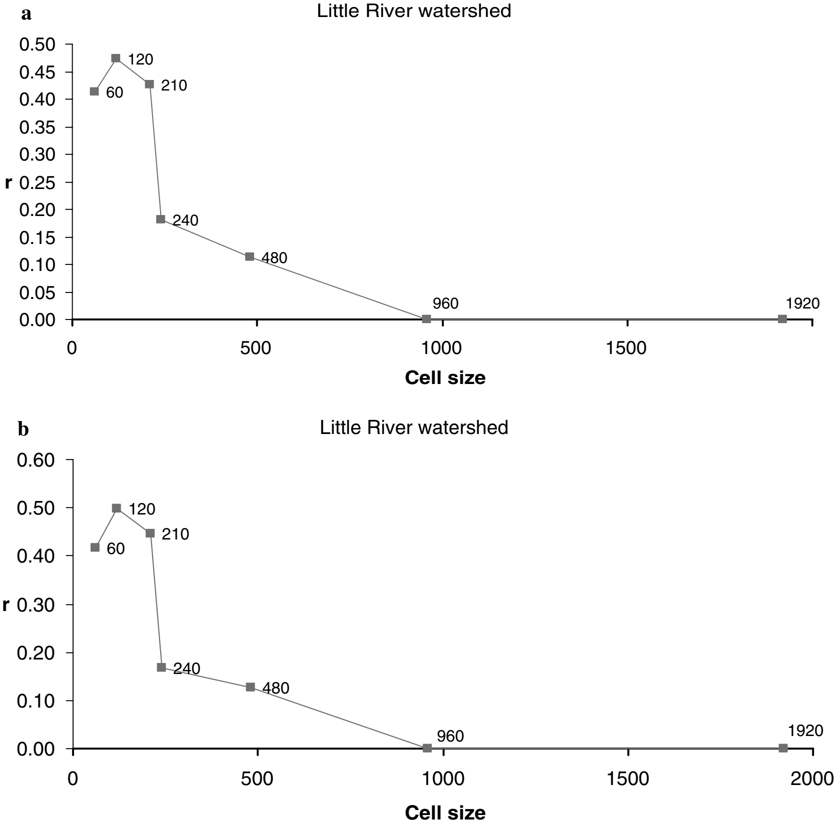


Fig. 9. Resample correlation coefficient (*r*) between model output of 30-m cell size compared to other cell sizes for Little River of: **a** soluble nitrogen content (ppm), and **b** soluble phosphorous content (ppm)

However, our work is of much higher resolution and only after gross resampling of 30-m data to 60, 120, 210, etc) do our datasets reach the critical threshold of 0.01 as described by Garbrecht and Martz (1994). Thus, our findings apply more broadly to high-resolution (3- and 30-m) cells in large watersheds averaging approximately 35,000 ha. In addition, our results are consistent with Vieux and Needham (1993) results showing the variance in yield that they experienced is similar in magnitude to the variance we show when resampling land cover data to lower resolution cell sizes. Our results uniquely show the effects of spatial resolution by using two independent collections of data, at 3 m and 30 m for both elevation and land cover.

9 Conclusions

GIS provide excellent data handling capabilities for the development of databases appropriate for watershed and water-quality modeling. We developed databases of elevation, land cover, and soils at various resolutions in four different watersheds. The database development involved significant processing but resulted in datasets from which parameters for AGNPS were generated directly. The datasets and derived parameters provide a basis for analyzing the effects of both resolution and resampling. Techniques used included comparison of values at specific points across multiple datasets, comparison of total areas, and statistical analysis using correlation methods.

Results of this analysis indicate that elevation values at specific points compare favorably ($r = 0.90$) between 3- and 30-m post spacing datasets. When the 30-m data are resampled to 60, 120, 210, 240, and 480 m, a single profile across the watershed retains its shape, but at resampling to 960 and 1,920 m, the profile ceases to be the same. Comparison at random sample points of specific elevation values across various resamplings shows a continual degradation in correspondence to the original 30-m elevations.

Resolution and resampling cause significant changes in land cover values, perhaps because of the categorical scaling of the data. Comparison of independent collections of the same land cover categories at 3 and 30 m show significant differences (pseudo $R^2 = 0.14$) in assigned covers at random test points. Resampling to coarser resolutions also degrades land cover, and there is a significant difference (Table 4) in values at sample points. Similarly, total area calculations for each land cover in a particular watershed show significant differences (Table 5), indicating that the resampling is degrading the original data.

Certain derived parameters, such as land slope, parallel the results of the GIS datasets that are their base. Degradation or change of values increases as resolution becomes lower and more cells are combined to create a single output value. We also examined AGNPS model output for nitrogen and phosphorous. This examination showed that the quality of AGNPS model output, soluble nitrogen and phosphorous, also degrade with resolution. Further analysis of all model output is underway.

As a general conclusion, our analysis shows that data should be collected at the desired resolution since resampling, even for continuous datasets, will degrade or modify the original data values. This result holds true for originally collected data, such as elevation and land cover, derived data, such as land slope and flow direction, and model outputs, nitrogen and phosphorous.

Original results for this work include the examination of resolution effects for independent collection of elevation and land cover data. Elevation compares well across resolutions of 3 m and 30 m with 90 % correlation. Land cover does not compare well. These results may reflect our abilities to create better continuous surfaces (elevation) at different resolutions than discrete surfaces (land cover). This work indicates a threshold between 480 m and 960 m cells at which resampling completely degrades the data for use in the AGNPS model. This threshold may result from basic model assumptions used in AGNPS, which are sensitive to the area and to an aggregation spatial maximum.

References

Anderson JR, Hardy EE, Roach JT, and Witmer RE (1976) A Land Use and Land Cover Classification System for Use with Remote Sensor Data. *U.S. Geological Survey Professional Paper 964,* pp 28

Berndt MP, Hatzell HH, Crandall CA, Turtora M, Pittman JR, Oaksford ET (1998) Water Quality in the Georgia-Florida Coastal Plain, Georgia and Florida, 1992-96. *U.S. Geological Survey Circular 1151*

ERDAS Inc. (1999) *ERDAS field guide*, Fifth Edition: Atlanta, Ga

Finn MP, Usery EL, Scheidt DJ, Beard T, Ruhl S, Bearden M (2002) AGNPS Watershed Modeling with GIS Databases. *Proceedings 2nd Federal Interagency Hydrologic Modeling Conference*. Las Vegas, Nevada

Garbrecht J, Martz L, (1994) Grid Size Dependency of Parameters Extracted From Digital Elevation Models. *Computers & Geosciences* 20(1): 85–87

Gesch D, Oimoen M, Greenlee S, Nelson C, Steuck M, Tyler D (2002) The National Elevation Dataset. *Photogrammetric Engineering and Remote Sensing* 68(1): 5–11

Gilliom RJ, Mueller DK, Zogorski JS, Ryker SJ (2002) A National Look at Water Quality. *Water Resources IMPACT*, Vol 4, No 4 American Water Resources Association, Middleburg, Virginia

Hamilton PA (2002) Water-Quality Patterns in Some of the Nation's Major River Basins and Aquifers. *Water Resources IMPACT*, Vol 4, No 4 American Water Resources Association, Middleburg, Virginia

Hodgson MA (1995) What Cell Size Does the Computed Slope/ Aspect Angle Represent? *Photogrammetric Engineering and Remote Sensing* 61(5): 513–517

Inskeep WP, Wraith JM, Wilson JP, Snyder RD, Macur RE, Gaber HM (1996) Input Parameter and Model Resolution Effects on Predictions of Solute Transport. *Journal of Environmental Quality* 25: 453–462

Leavesley GH, Hay LE, Viger RJ, Markstrom SL (2002) *Use of a Priori Parameter-Estimation Methods to Constrain Calibration of Distributed-Parameter Models*. In: Duan Q, Gupta HV, Sorooshian S, Rousseau AN, Turcotte R (eds) *Calibration of Watershed Models*. Water Science and Application, vol 6, American Geophysical Union, Washington, DC, pp 255–266

Long JS (1997) *Regression Models for Categorical and Limited Dependent Variables*. Sage Publications, Thousand Oaks, California

Olivieri LJ, Schall GM, Logan TJ, Elliot WJ, Motch B (1991) Generating AGNPS Input Using Remote Sensing and GIS. *International Winter Meeting, American Society of Agricultural Engineers*, Chicago, Illinois

Tim US, Mostaghimi S, Shanholtz VO (1992) Identification of Critical Non-Point Pollution Source Areas Using Geographic Information Systems and Water Quality Modeling. *Water Resources Bulletin* 28 (5): 877–887

USGS (2002a) *National Elevation Dataset*. Internet at: http://gisdata.usgs.net/ned/PaperNED.htm

USGS (2002b) *National Land Cover Characterization Project*. Internet at: http://landcover.usgs.gov/nationallandcover.html

Vieux, BE (1993) DEM Aggregation and Smoothing Effects on Surface Runoff Modeling. *Journal of Computing in Civil Engineering* 7 (3): 310–338

Vieux BE, Needham S (1993) Nonpoint-Pollution Model Sensitivity to Grid-Cell Size. *Journal of Water Resources Planning and Management* 119 (2): 141–157

Viger RJ, Markstrom SL, Leavesley GH (2002) The GIS Weasel: An Interface for the Development of Spatial Parameters for Physical Process Modeling. *Proceedings 2nd Federal Interagency Hydrologic Modeling Conference*. Las Vegas, Nevada

Vogelmann JE, Howard SM, Yang L, Larson CR, Wylie BK, and Van Driel N (2001) Completion of the 1990s National Land Cover Data Set for the Conterminous United States from Landsat Thematic Mapper Data and Ancillary Data Sources. *Photogrammetric Engineering and Remote Sensing* 67 (6): 650–662

Witte J, Theurer FD, Baker KD (1995) *AGNPS Version 5.00 Verification: Software*. Internet at: http://www.sedlab.olemiss.edu/AGNPS.html

Wolock, DM, Price CV (1994) Effects of digital elevation model map scale and data resolution on a topography-based watershed model. *Water Resources Research* 30 (11): 3041–3052

Young RA, Onstad CA, Bosch DD, Anderson WP (1994) AGricultural Non-Point Source Pollution Model, Version 4.03, *AGNPS User's Guide*. North Central Soil Conservation Research Laboratory, Morris, Minnesota

Young RA, Onstad CA, Bosch DD, Anderson WP (1995) *AGNPS: An Agricultural NonPoint Source Model*. In: Singh VP (ed) *Computer Models of Watershed Hydrology*. Water Resources Publications, Highlands Ranch, Colorado pp 1011–1020